

周报

- 1.本周前三天主要在准备组会论文报告
 - 2.其余时在完成 Simhash 相似性搜寻和聚类的代码设计、测试与实现。
- 依据上周的设想,主要设计了两个类, VolumeSplitter 与 SimhashCluster, 头文件分别如图 1 和图 2 所示

```
struct int3
{
    int x;
    int y;
    int z;
};

class VolumeSplitter
{
public:
    VolumeSplitter(int innumBlock);

    void SetVolumeSize(const int inx, const int iny, const int inz);
    void SetSplitBlockSize(const int inx, const int iny, const int inz);
    void SetDataArrayElement(const float value, const int elementIndex);

    void GetBlockData(const int inblockID, unsigned char * outDataArray);
    int3 GetVolumeSize() const;
    int3 GetSplitBlockSize() const;
    float GetDataArrayElement(const int elementIndex) const;

    bool AllocatingMemory();
    bool LoadVolumeDataAsBlocks(const std::string rawFilePath); //load volume data in block style, in order of xyz firection.

    ~VolumeSplitter();

private:
    void Init(int innumBlock);
    void ReleaseMemory();

    unsigned char *_data; //data为原始数据
    int _numBlock;
    int3 _volumeSize;
    int3 _splitBlockSize;
    unsigned char *_indexData;
```

图 1 体数据分块读取操作

```

#ifndef _SIMHASH_CLUSTER_H_
#define _SIMHASH_CLUSTER_H_
#include <iostream>
#include <cstdio>

class SimHashCluster
{
public:
    SimHashCluster(int innumBlock);
    ~SimHashCluster();

    void SetFingerprint(const int inIndex, const unsigned char* fingerprintFromSimhash);
    void GetOneFingerprint(const int inIndex, unsigned char* outFingerprint);
    unsigned char * GetVolumeFingerprintSets();
    bool AllocatingMemory();

private:
    void ConstructFingerprintSetTables();//构建整个指纹集合查找表
    void Replace();//置换指纹中的不同部分
    void SortFingerprintSet();//对指纹表进行排序，得到有序的指纹表以便快速查找(optional)
    int GetHammingDistance(unsigned char* firstValue, unsigned char* secondValue, int sim_hash_size);//计算汉明距离
    void ReleaseMemory();

    int _hamming_distance_threshold;//用于判断是否属于一类的汉明距离阈值
    unsigned char * _volumeFingerprints;//大小为体数据block的个数*128位哈希值，128位=16*8bit;故大小为numBlock*16;
    static const int _hashSizeinUchar;
    int _numBlock;
};

#endif // _SIMHASH_CLUSTER_H_

```

图 2 相似哈希聚类

目前，VolumeSplitter 已实现、测试完毕，SimhashCluster 完成一半，使用 cluster 进行指纹集合的复制，移位置换，相似度查询，聚类等操作还没有实现。本周将实现并测试这部分内容，若完成，则压缩端全部完成，可进行压缩效率和质量的测试。然后进入绘制阶段。